University of Groningen

CLOCK Gene Variants Associate with Sleep Duration in Two Independent Populations

Allebrandt, Karla V.; Teder-Laving, Maris; Akyol, Mahmut; Pichler, Irene; Mueller-Myhsok, Bertram; Pramstaller, Peter; Merrow, Martha; Meitinger, Thomas; Metspalu, Andreas; Roenneberg, Till

Published in:
Biological Psychiatry

DOI:
10.1016/j.biopsych.2009.12.026

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Allebrandt, K. V., Teder-Laving, M., Akyol, M., Pichler, I., Mueller-Myhsok, B., Pramstaller, P., Merrow, M., Meitinger, T., Metspalu, A., Roenneberg, T., & Müller-Myhsok, B. (2010). CLOCK Gene Variants Associate with Sleep Duration in Two Independent Populations. *Biological Psychiatry*, *67*(11), 1040-1047. https://doi.org/10.1016/j.biopsych.2009.12.026

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

CLOCK Gene Variants Associate with Sleep Duration in Two Independent Populations

Karla V. Allebrandt, Maris Teder-Laving, Mahmut Akyol, Irene Pichler, Bertram Müller-Myhsok, Peter Pramstaller, Martha Merrow, Thomas Meitinger, Andreas Metspalu, and Till Roenneberg

Background: Sleep is an active and complex behavior, yet it has two straightforward properties—timing and duration. Clock genes are associated with dysfunctional timing of sleep, mood, and obesity disorders, which are commonly associated with sleep duration.

Methods: Sleep duration was assessed in Central Europe, Estonia, and South Tyrol ($n \approx 77{,}000$) with the Munich ChronoType Questionnaire. It showed a Gaussian distribution in all investigated populations after averaging over a standard workweek and normalization according to age and gender. A follow-up, two-stage design, linkage disequilibrium-based association study was conducted with subjects from South Tyrol (discovery sample; $n = 283$) and with short (< 7 hours) and long (> 8.5 hours) sleepers from Estonia (confirmation sample; $n = 1011$). One hundred ninety-four single nucleotide polymorphism markers covering 19 candidate clock genes were genotyped in the discovery sample, and two of the best association signals (analyzed by a linear regression model) were investigated in the confirmation sample.

Results: Single and multi-marker associations were found within a *CLOCK* gene intronic region (rs12649507 and rs11932595). In a meta-analysis between South Tyrol and Estonia association signals, rs12649507 ($p = .0087$) remained significant. Significance persisted only for the multiple-marker association signal of the rs12649507/rs11932595 haplotype *GGAA* with long sleep ($p = .0015$).

Conclusions: We report an association between variants of the human *CLOCK* gene and sleep duration in two independent populations. This adds another putative function for *CLOCK* besides its possible involvement in circadian timing, depression, obesity, and personality.

Key Words: *CLOCK*, sleep duration, clock genes, MCTQ, short sleepers, long sleepers

Sleep is one of the most conspicuous aspects of our existence, drastically changing our state of consciousness for approximately one-third of our lives. The question "why do we sleep?" is one of the hot topics in biology. On the one hand, even small doses of sleep deprivation cause decrements in performance, and complete sleep deprivation is fatal in experimental animals; but on the other hand, pharmacological agents such as modofanil can be used to allay sleep with remarkably small side effects. Hypotheses regarding why we sleep concern neuronal reorganization and repair (1–4). We remain, however, far from fully understanding the function of sleep.

Despite its complexity, sleep has two straightforward characteristics—timing and duration. They are influenced by two physiological mechanisms: a homeostatic (how long have we been awake) and a circadian process (5,6). The latter refers to the circadian clock, an endogenous mechanism that specifies a temporal window—a phase—within the day for a myriad of processes, from gene expression to behavior, including sleep. The timing of sleep is often used to assess entrainment of the circadian clock. This entrainment is specific to individuals, yielding different so-called chronotypes (7).

Sleep timing can be distinguished from sleep duration. There are just as many short and long sleepers among early chronotypes as there are among late ones (7). Sleep duration has been associated with disease and mortality (8); personality traits, such as self-criticism, emotional reliance on another person, or agreeableness (9,10); and learning and memory (11). Lack of sleep has been associated with obesity and diabetes (12). The effect of sleep deprivation on performance shows large differences among self-reported short (<6 hours) versus long sleepers (>9 hours); short sleepers apparently tolerate a higher homeostatic sleep pressure than long sleepers (13), presumably because a genetically regulated mechanism allows them to recover sleep homeostasis faster or to better withstand sleep pressure relative to long sleepers. Other sleep qualities, besides timing, have been associated with so-called clock genes (those involved in generating the circadian rhythm at the cellular level). Sleep structure and homeostasis, for example, correlate with *PER3* gene variants (14), and clock gene (*PER1* and *PER2*) expression correlates with sleep deprivation and recovery phase in mice (15). The temporal regulation of non-rapid eye movement sleep is compromised in NPAS2 −/− mice (16).

Sleep duration has a genetic component corresponding to 40% heritability, as shown in twin studies (17), but its genetics will surely not be simple (18). Timing and duration of sleep are influenced by the circadian clock, as indicated by functional studies on model organisms (19–22) and genetic association studies (for review see 23,24). Associations between sleep duration and gene variants in general human populations have so far been addressed only once by a small-scale genome-wide study (25).

From the Centre for Chronobiology (KVA, TR), Ludwig Maximilians University of Munich, Munich, Germany; Estonian Biocentre (MT-L, AM); Estonian Genome Center and IMCB of University of Tartu (AM), Tartu, Estonia; Institute of Human Genetics (KVA, MA, TM), Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg; Institute of Genetic Medicine (IP, PP), European Academy Bozen/Bolzano (EURAC), Bolzano, Italy—Affiliated Institute of the University of Lübeck; Max-Planck-Institute of Psychiatry (BM-M); Institute of Human Genetics (TM), Klinikum rechts der Isar, Technische Universität München, Munich; Department of Neurology (PP), University of Lübeck, Lübeck, Germany; Department of Neurology (PP), General Central Hospital, Bolzano, Italy; and the Department of Chronobiology (MM), University of Groningen, Haren, The Netherlands.

Address correspondence to Till Roenneberg, B.Sc., Ph.D., Medical Faculty, Ludwig-Maximilians-University, Goethestr. 31, D-80,336 München, Germany. E-mail: roenneberg@lmu.de.

Received Aug 12, 2009; revised Nov 5, 2009; accepted Dec 11, 2009.

0006-3223/$36.00
doi:10.1016/j.biopsych.2009.12.026

© 2010 Society of Biological Psychiatry

Although the genetics of sleep duration has been addressed in humans, associations between clock genes and sleep duration have never been investigated systematically. We therefore chose 19 clock genes as candidates and investigated a set of high-density markers in a two-stage design association study, involving populations from South Tyrol (discovery sample) and Estonia (confirmation sample). Single nucleotide polymorphism markers (SNPs) were selected on the basis of a linkage disequilibrium (LD) strategy or on their possible relevance as nonsynonymous or splicing region variants. Sleep duration depends both on age and on gender. We therefore used our Central European and Estonian databases ($n \approx 65{,}000$ and $n \approx 12{,}000$, respectively) to establish algorithms describing the inherent and systematic relationships. These were then used to normalize the phenotype of sleep duration (averaged over a standard week).

Here, we report the association between sequence variants in the human *CLOCK* gene and sleep duration in two independent European populations. Because we have investigated general populations, it is likely that the influence of the variability of this gene on the distribution of sleep duration is not an isolated case and might apply to other populations.

Methods and Materials

Subjects

All subjects were selected from ethnically homogeneous, Caucasian populations. The local ethic committees approved all subjective assessments and genotyping analyses. The discovery sample was recruited from healthy individuals—participants of the MICROS study (genetic study of three population microisolates in South Tyro) (26)—living in one of three valleys in South Tyrol (Stilfs, Martell, and Langtaufers; all situated within a German-speaking region of Northern Italy, of approximately 1400 km^2). Three hundred thirty-three randomly selected individuals (total population size ≈ 2600) were phenotyped and genotyped. Ages ranged from 15 to 75 years with a slight majority of women (178 women vs. 155 men). Population (sub)structure and consanguinity between study participants were estimated on the basis of a genome-wide dataset from the same population (ongoing study) with mixed model and regression (27). Population stratification estimates confirmed a reduced common genetic background among study participants ($\lambda = 1.009$), thus, eliminating the need for family-based association analysis in this study.

Samples for the replication study were obtained through the Estonian genome project (http://www.geenivaramu.ee) and were recruited from nonrelated Estonians. In this case, there were 1011 subjects between the ages of 18 and 69 years, again with a surplus of women (683 women and 328 men). The Estonian sample was characterized with respect to other European populations with the principal components analyses method (28). The DNA samples chosen for genotyping were drawn from the extremes of a pool of 5098 subjects in the distribution of normalized average sleep duration (they were part of a growing Estonian database that meanwhile consists of approximately 12,000 subjects). The phenotype was assessed with the Munich ChronoType Questionnaire (MCTQ) (see following text). Exclusion criteria for both South Tyrolean and Estonian samples were: 1) use of an alarm clock on free days; 2) shift-work during the last 3 months; and 3) extreme outliers of the normal distribution (e.g., subjects sleeping <4 hours).

Phenotyping

Average sleep duration was assessed with the short version of the MCTQ (29). Sleep-onset was calculated by adding sleep latency to the time of sleep preparation, and sleep duration was calculated by subtracting sleep onset from sleep end. All values were assessed separately for workdays and for free days (SD_w and SD_f, respectively). Subjects were also asked how many days/week they worked (N_{wd}). Average sleep duration (SD_{av}) was calculated by the following equation:

$$SD_{av} = (N_{wd} \times SD_w + [7 - N_{wd}] \times SD_f)/7$$

The SD_{av}, was then corrected for its dependencies on age and gender, yielding the final quantitative phenotype: SD_{asc}. Algorithms for the normalization of age- and gender-dependent changes were based on two databases comprising more than 65,000 subjects from Central Europe (mainly Germans but also Austrians, Dutch, and Swiss) and the Estonian database ($n \approx 12{,}000$; Figures 1A and 1B). The equations of the fits (Table S1 in Supplement 1) were used to normalize sleep duration to a theoretical "neutral" 30 years old.

Marker Selection

We screened 19 clock genes, distributed across 10 chromosomes (Figure 2). Because most of these genes still await validation as clock genes in humans, coding sequence homologies of animal orthologs (e.g., the *Drosophila* gene *Jetlag* corresponds to the human gene *FBXL15*) were determined with the help of the BLAST program (http://www.ncbi.nlm.nih.gov/). One hundred ninety-four SNPs (Table S2 in Supplement 1) across the 19 genes were selected from three sources. An LD-based approach with the Caucasian (CEU) HapMap Phase II panel produced 155 tags (criteria were minor allele frequency [MAF] > 20% and $r^2 > .8$); the National Center for Biotechnology Information databank gave rise to 24 nonsynonymous variants; 5 splice region variants were selected from the Ensembl genome browser. The number of markers selected/gene region (including up and downstream regions) depended mostly on the LD of these markers with surrounding variants (tagging with r^2 cutoff = .8), assessed with the software Haploview (30) as predicted by the Tagger-pairwise Tagging algorithm. Redundant tags replaced tag SNPs that did not work with the iPLEX assay (Sequenom, San Diego, California). Nonsynonymous or splicing region variants were not necessarily polymorphisms and were selected independently of the LD-based approach.

Genotyping

The DNA was extracted from ethylenediaminetetraacetic acid-treated blood samples with a salting out procedure (31) and quantified with a NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, Massachusetts). Discovery samples were genotyped by polymerase chain reaction (PCR) primer extension assay with a Sequenom iPLEX genotyping platform, following the manufacturer's instructions. After amplification of target sequences and before the primer extension reaction of the iPLEX protocol, the DNA amplification quality (24 wells/plate) was verified in an agarose gel. Base-extension products were detected in a MassARRAY time-of-flight mass spectrometry system (32). Replication samples were genotyped by using Taqman SNP Genotyping Assays on an Applied Biosystems (Foster City, California) 7900HT Fast Real-Time PCR machine (33). The PCR reactions contained 20 μL (10 ng) of DNA, 10 μL of TaqMan

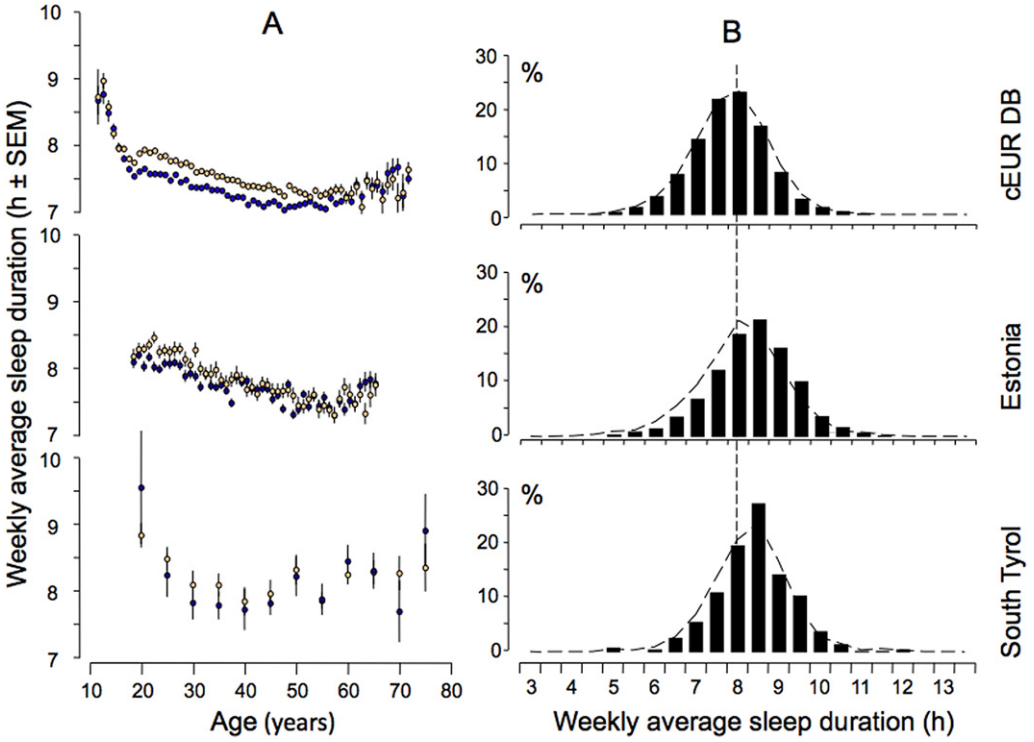

Figure 1. Sleep duration depends on age and gender. **(A)** On average, women (open circles) sleep longer than men (filled circles). The age-dependency of not-normalized averaged weekly sleep duration is plotted, separately for the genders, for the large Central European database (cEUR DB) (upper left panel) as well as the Estonian (middle left panel) sample annually and South Tyrolean (bottom left panel) sample on a 5-year basis. For details see text. **(B)** The distributions of weekly average sleep duration in the different populations. The cEUR DB (upper right panel), Estonia (middle right panel), and South Tyrol (bottom right panel). The distributions of the raw data are drawn as stippled curve, and the distribution of the age and gender corrected data are drawn as bars. For orientation purposes, a vertical line is drawn through sleep duration of 8 hours.

(Applied Biosystems) Genotyping Master Mix (2×), and 1 μL of Taqman Pre-Designed SNP Genotyping Assay (20×). Samples were denatured at 95°C for 10 min followed by 40 cycles at 92°C for 15 sec and 60°C for 1 min. Genotyping was performed in two genotyping facilities, the Helmholtz Zentrum München and the Estonian Biocentre.

Statistical Analysis

Quantitative trait association was tested with linear regression model analysis for association of normalized sleep duration with the genotypes with the PLINK (http://pngu.mgh.harvard.edu/purcell/plink/) package version 1.04 (34). Multi-marker analysis was performed with the haplotype-based quantitative trait association test from the PLINK package with a sliding window of two SNPs. A meta-analysis was conducted with the *R* package rmeta version 2.15 to estimate the magnitude of the combined effect of the results from both samples.

Results

Phenotyping Sleep Duration

Optimal phenotyping is an important prerequisite for achieving unequivocal results in phene-gene association studies. Similar to sleep timing (35), sleep duration depends systematically on both age and gender (Figure 1). Access to large databases on sleep behavior allowed us to establish equations (Table S1 in Supplement 1) that were used for normalizing sleep duration with regard to the influences of age and gender (adults in the Central European sample: women, $n = 30{,}959$, $r = .87$; men, $n = 28{,}236$, $r = .87$; adults in the Estonian sample: women, $n = 8035$, $r = .86$; men $n = 3676$, $r = .73$). Additional fits for both young women and men (ages 10–20) were possible for the large Central European sample ($n = 5111$ $r = .93$ and $n = 3904$, $r = .96$, respectively). The relatively small South Tyrolean sample did not allow significant estimations (women: $n = 178$, $r = .14$; men: $n = 155$, $r = .13$). We applied the correction algorithms established in the Central European database to the South Tyrolean data, on the basis of the high correlation of the age- and gender-dependent changes in Central Europe and Estonia (women: $r = .9$; slope = 1.3; men: $r = .8$; slope = 1.1).

Figure 1B shows the distributions of the averaged weekly sleep duration for the Central European, Estonian, and Tyrolean populations before and after normalization. After normalization SD_{asc} distribution and mean were not different between the Estonian and Tyrolean populations ($n = 5098$, $SD_{asc}Ø = 8.07$, $SD_{asc}Ø = 8.14$, respectively; *t* test: $p > .1$; analysis of variance $p > .05$, Figure 1B). Before normalization, they were significantly different ($p < .0001$).

For the replication study, 507 short (<7 hours) and 504 long (>8.5 hours) sleepers were selected from the distribution of normalized average sleep duration in the Estonian database (Figure 1B). This strategy is powerful for association studies, because genetic differences between the extremes of a quantitative trait distribution should become more evident (36). We have not selected the extremes from the total pool of phenotyped subjects used for normalization (approximately 12,000 subjects, Figure 1), because by the time genotyping started only a subset of the DNA samples were available.

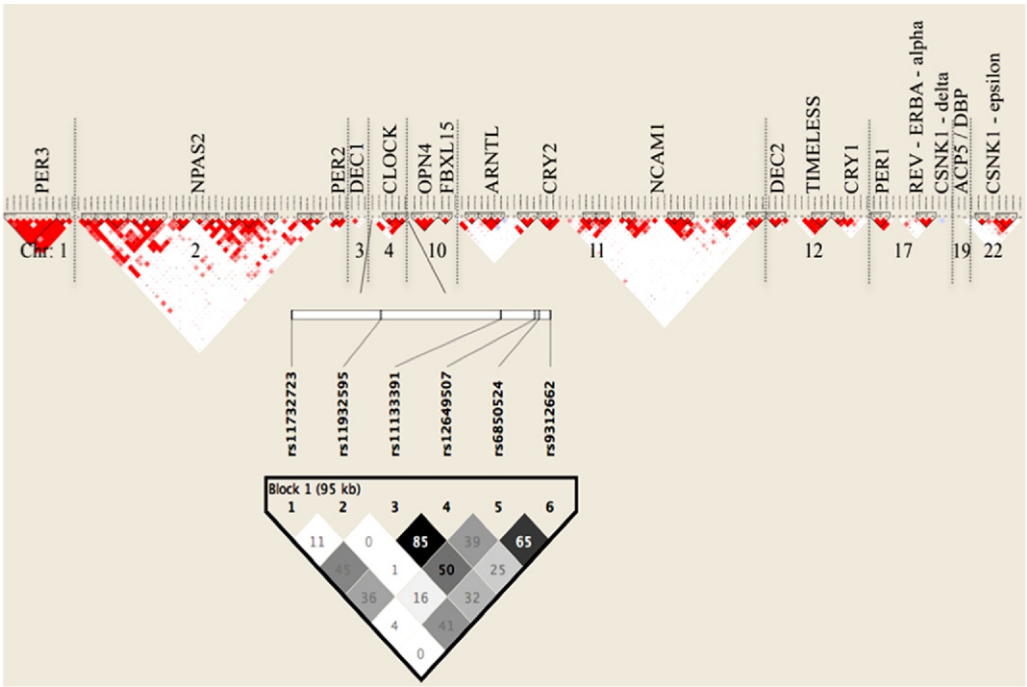

Figure 2. Linkage disequilibrium structure (triangle plots) for all investigated single nucleotide polymorphism markers (SNPs) (rs numbers below gene names). Dark diamonds indicate strong historical linkage disequilibrium (LD), *D′*-based haplotype boundaries, among SNPs. The amplified image is an *r*2-based triangle plot of the *CLOCK* gene with genotyped tags indicating, from black to white, complete to absent LD on the basis of HapMap data. Chr, chromosomes.

Genotyping Efficiency

Total call rate in the discovery sample was .94. Fifty of the 333 individuals were removed for low genotyping (call rate/individual was between 80% and 89%), because a large number of markers were genotyped. All SNPs included in the analysis had an MAF >.01 and total call rate of .99. Two SNPs were genotyped in the confirmation sample (Estonia; $n = 1011$) with a call rate of 100% and Hardy–Weinberg-equilibrium (HWE) ($p > .05$) in both samples. All SNPs with a call rate under 85% were excluded from the analysis (18 of the 194 investigated variants). Nineteen monomorphic variants and 5 variants that were not in HWE (cutoff: $p < .01$) were excluded from the association analysis. Only two cases of heterozygote underrepresentation were present among investigated SNPs, and those were not in HWE, as expected. All other functional SNPs (except monomorphic ones) had a higher proportion of heterozygotes in relation to the homozygotes. The average allele frequency was .328 ± .110. The MAFs for rs12649507 and rs11932595 in the confirmation sample were .30 and .49, respectively. Marker names, genomic location, gene, function, alleles, HWE, and minor frequencies are listed in Table S2 in Supplement 1.

Association Analysis

The most significant phenotype-genotype associations in the discovery sample (from South Tyrol) were found for SNPs located in the intronic region of the *CLOCK* gene (rs12649507, $p = .0051$; rs11932595, $p = .0080$; Figure 3, Table 1). These alleles were not in LD with each other. The fact that we have removed 50 of the subjects of the discovery sample to achieve a total call rate of 99% did not modify the final results. When including all subjects for the analysis, the two reported associations were still the most significant (rs12649507, $p = .0059$; rs11932595, $p = .0021$). From the six markers tagging the *CLOCK* gene (Figure 2), only five were considered for the association analysis (rs6850524, rs9312662, rs11732723, rs11932595, and rs12649507)—one of them, rs111333911, was eliminated on the basis of a low call rate. The number of investigated SNPs/gene did not bias the discovered associations. Other clock genes investigated in this study had many more variants than the *CLOCK* gene (Figure 2). The SNPs in four other genes (*NPAS2*, *Timeless*, *CSNK1D*, and *Per3*) produced *p* values < .05 for the association with SD_{asc} (Table 1) but were above the threshold of $p < .01$ that we had set for the second stage of this study. Neither of the aforementioned results (the associations in *CLOCK* nor those in the four additional genes) remained significant after

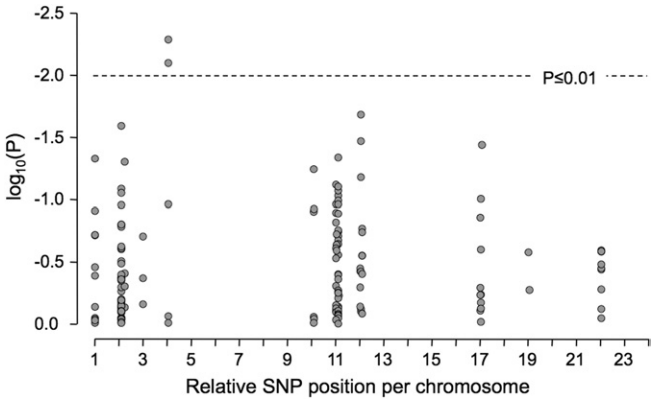

Figure 3. Significances (plotted as $-\log_{10}$ of nominal *p*) resulting from the associations between the investigated single nucleotide polymorphism markers (SNPs) and normalized sleep duration in the discovery sample (South Tyrol). The x-axis represents the relative genomic position/chromosome, and the y-axis represents the distribution of the nominal $-\log_{10}(p)$ values in the discovery sample. The two SNPs, which fulfill the inclusion criterion for replication ($p < .01$; above stippled line), map to the *CLOCK* gene located on chromosome 4.

Table 1. Associations of Single and Multi-Markers with Normalized Average Sleep Duration

Single Marker Associations			Discovery Sample		Confirmation Sample		Meta-Analysis
Genes	SNP rs	Location	*n*	*p*/Effect Allele	*n*	*p*/Effect Allele	*p*
CLOCK	rs12649507	Intron 1	282	.0051/A	1011	.0453/A	.0087
CLOCK	rs11932595	Intron 11	275	.0080/G	1011	.0472/A	.2443
NPAS2	rs7598826	Intron 1	281	.0255/A	—	—	—
TIMELESS	rs2291738	Intron 22 - boundary	281	.0207/G	—	—	—
CSNK1D	rs7209167	Intron 1	277	.0363/G	—	—	—
PER3	rs10462021	Exon 20	268	.0468/G	—	—	—

Multiple-Markers Associations							
Genes	SNP rs	Effect Alleles	*n*	*p*/Effect Allele	*n*	*p*/Effect Allele	*p*
CLOCK	rs11932595, rs12649507	*GG*	274	.0003	1011	.9022	.376
		AA		.0005		.8963	.377
		GA		.4103		.0019	.0015
		AG		.7995		.0316	.0601
CLOCK	rs11732723, rs11932595	*GG*	275	.0008	—	—	—
		AA		.0862			
		GA		.0997			
		AG		.6185			
TIMELESS	rs10783783, rs2291739	*TC*	257	.0008	—	—	—
		TT		.1397			
		CT		.3784			
		CC		.4096			

Gene name, single nucleotide polymorphism marker (SNP) rs code, location, sample size, and nominal *p* values/effect alleles are listed for the discovery and replication samples. Dashed lines mean that the respective SNPs have not been genotyped in the confirmation sample.

Bonferroni correction for multiple comparisons. Assuming that a two-stage design study would be more meaningful than an empiric correction, we tried to replicate our findings for the best SNPs (rs12649507 and rs11932595; α level $p < .01$) in an independent sample from our Estonian database. Indeed, significant associations were found for rs12649507 and rs11932595 ($p < .05$; Table 1). However, only the rs12649507 association could be replicated, whereas a heterogeneous effect was observed for rs11932595 in comparison with the discovery sample. Thereby, we conducted a meta-analysis of the results of the individual studies; the rs12649507 association signal remained significant ($p = .0087$), whereas rs11932595 association lost significance ($p = .2443$).

We also performed a two-marker haplotype analysis for all possible allelic combinations for all investigated SNPs. The same *CLOCK* SNPs, which produced the best results for the single marker analysis, also produced the most significant haplotype associations in the discovery sample (lowest $p = .0003$). An association signal was also found in the confirmation sample (lowest $p = .0019$; Table 1), although heterogeneous association with sleep duration length was observed in the two samples. A meta-analysis of the results from both samples validated only the association of rs12649507/rs11932595 *GGAA* with long sleep ($p = .0015$). The averaged SD_{asc}/genotypes and haplotypes for rs12649507 and rs11932595 are presented in Supplement 1 (Supplemental Results) and on Figure 4, respectively. In the discovery sample, two other association signals were found in the multi-marker analysis at the α level of significance of $p < .01$ (Table 1). One of them was also a haplotype of *CLOCK* gene markers (rs11732723/rs11932595; $p < .001$), but significance at the single SNP level for rs11732723 was not observed.

Discussion

The distribution of weekly averaged sleep duration depends on age and gender in Central Europe and Estonia (Figures 1A and 1B). The most drastic shortening in sleep duration (by approximately 2 hours) occurs between childhood and the end of puberty. From that age until approximately the age of menopause (at approximately the age of 50), men sleep on average (15 min) less than women, in accordance with earlier studies (37). This difference has been associated with concentration changes of sex hormones (38). Sleep duration decreases steadily in both genders from the end of adolescence (at approximately the age of 20) to the age of menopause (39). From that age on, the gender differences disappear, and sleep duration increases slightly with age (Figures 1A and 1B). The confounding influences of these systematic dependencies were eliminated by our normalization procedure to optimize phenotyping for the genetic associations described here. Due to the limited sample size of the South Tyrolean database (Figure 1A), normalization for this discovery sample was based on the Central European database.

Despite the large variables set (152 variants distributed over 19 genes) investigated in the discovery sample, the top two most significant signals were located within the same gene, the *CLOCK* gene—an essential transcription factor of the molecular circadian clock. *CLOCK* encodes a basic helix-loop-helix–PAS (a domain named after three proteins it occurs in: Per, Arnt, and Sim) transcription factor (40), which forms a complex with the protein BMAL1. This heterodimer is the activating component of a negative feedback-loop that is critically involved in circadian rhythm generation at the cellular/molecular level (41,42). In addition to its role within the circadian system, functional studies showed CLOCK to be also involved in sleep regulation (21).

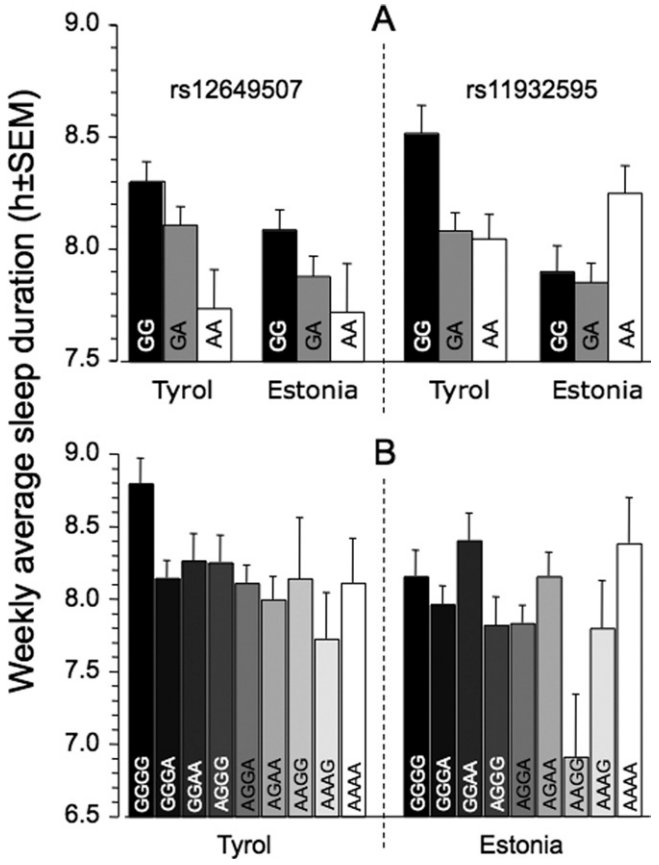

Figure 4. Normalized average sleep duration among subjects having different genotypes **(A)** and haplotypes **(B)** for *CLOCK* single nucleotide polymorphism markers (rs12649507/rs11932595) associated with sleep duration in the discovery (South Tyrol) and confirmation (Estonia) samples.

In this study, we have found the association of the *CLOCK* gene variants with sleep duration in two independent populations, indicating that sleep duration has a genetic component related to the circadian mechanism. The *CLOCK* markers investigated in this study cosegregate with 59 other polymorphic sites of this gene (CEU HapMap data), mostly located in intronic region. Between the analyzed *CLOCK* SNPs, one (rs9312662) was a marker for a polymorphic synonymous coding variant of *CLOCK* (rs3736544) and for several intron boundary variants, but significant association of this marker with sleep duration was not found in this study. The *CLOCK* gene nonsynonymous exonic variants described until now are mostly rare mutations and cannot be responsible for a quantitative phenotype such as sleep duration. Short (<7 hours) and long sleepers (>8.5 hours) comprise approximately one-third in each of the investigated populations (Figure 1B). We are thus searching for frequent polymorphisms that could be associated with the relatively frequent extreme cases of sleep duration. If low frequency alleles would be contributing to this phenotype, we would rather expect them to be at coding or splicing region, so most of those have been selected in addition to the tagSNPs.

The *CLOCK* variant rs12649507 association signal was unidirectional in both populations and remained significant in a meta-analysis combining the independent results (Table 1). This polymorphism is located in intron one of the *CLOCK* gene, a region of limited variation (Figure S1 in Supplement 1). It cosegregates (r^2 >.80) with several neighboring variants (Table S3 in Supplement 1), covering a region of approximately 112 kb of the chromosome 4, including the *CLOCK* gene 5′ flanking region. The association of this SNP (rs12649507) showed the same tendency in both samples (i.e., sleep duration was decreased in subjects carrying the "*A*" allele). This was not the case for the SNP rs11932595, because "AA" homozygous for this polymorphism showed the shortest sleep duration in the South Tyrolean and the longest in the Estonian population (association signal that lost significance in the meta-analysis). This heterogeneous effect could have at least three reasons. The first reason is the different levels of LD of this marker with the causal allele. This SNP rs11932595 is located within a recombination region (Figure S1 in Supplement 1), subjected to fluctuations on LD levels in recently split populations (43). Allele frequencies for rs11932595 were different (p = .0320) between the discovery (A/G = .54/.46) and confirmation samples (A/G = .49/.51). The discovery sample had relatively more *AA* "homozygous, whereas the confirmation sample had more *GG*" homozygous, being homozygous frequencies significantly different (p = .0346). The second reason is different genetic environment. The extent of the contribution of other variants, implicated in this polygenic inheritance, plays a critical role in shaping the phenotype distribution. The third reason is phenotypic heterogeneity. The number of subjects assessed in the discovery sample was not representative for a quantitative phenotype distribution such as sleep duration. Only the confirmation sample haplotype association signal held its significance in the meta-analysis (Table 1). Indeed, the confirmation sample should give us a more precise estimate of the associations found than the discovery sample, because: 1) a sample comprising subjects with extreme sleep duration should increase the power to find genetic associations with this phenotype, because the larger effect observed on the tails of the distribution should reflect the contribution of several alleles from many quantitative trait loci; and 2) the sample size of the confirmation sample is three times larger than the discovery sample.

This is the first report that associates variants of the *CLOCK* gene with average sleep duration in human populations. An earlier study showed that a deletion of *CLOCK*'s exon 19 (due to a point mutation) shortened both the duration of sleep (by 1–2 hours) and that of rapid eye movement phases (21). Concerning other aspects of sleep behavior in human populations, a *CLOCK* gene polymorphism was associated with insomnia, an apparent inability to consolidate sleep (*T3111C*; rs1801260) located in the 3′ flanking region of the *CLOCK* gene (44). It was also associated with temporal sleep preferences, so-called morningness and eveningness (45,46). We have not included this SNP in our analysis, because its genotyping call rate was too low; it cosegregates with another SNP (rs7665846) that was included (Table S3 in Supplement 1). However, significant associations of rs7665846 were not found in this study, either for sleep duration or for sleep timing. Candidate gene approaches have yielded mixed information concerning *CLOCK* and *PER3*, with some studies reporting associations of SNPs or polymorphisms with sleep timing in a general population and others failing to confirm this, as reviewed elsewhere (23). The instrument used for phenotyping the subjects in these studies was a questionnaire that sought to identify early and late types as opposed to an instrument that would indicate the actual timing of sleep. Without taking sleep duration into account, a long sleeper could be classified as an early and/or a late sleeper, depending on whether onset of sleep or wake time is considered. Similar misinterpretations would apply to short sleepers. Therefore,

associations of those early and late types with *CLOCK* and *PER3* polymorphisms could have been biased by sleep duration.

Our top hit (rs12649507) cosegregates with *a CLOCK* variant (rs6843722) that has been associated with obesity in candidate genes studies (47,48). Although obesity is known to correlate with sleep duration (49), we have not found an association of rs12649507 with body mass index. This is not surprising, because there were only extremely weak correlations between normalized body mass index (Figure S2 in Supplement 1) and normalized sleep duration in our discovery ($r = .0029$) and confirmation samples (short SD, $r = .0010$; long SD, $r = .0141$). Another SNP (rs6850524) that is associated both with obesity (47) and with bipolar disorder (50) was investigated in the first stage of this study (South Tyrolean sample). Again, we found no association of sleep duration with it or other *CLOCK* SNPs in the discovery and confirmation samples. Because bipolar disorder is—like in the case of obesity—not independent of sleep duration (patients often have a decreased sleep need), a systematic picture might emerge: *CLOCK*'s association with different pathologies involves effects on sleep duration. However, this remains speculative, because causal links between these disturbances and sleep regulation are still controversial (51,52).

The function of CLOCK as a transcription factor or as a histone acetyltransferase implies that the genetic variability of this gene could influence its function on gene transcriptional activation. Environmental influences could also regulate CLOCK function as a histone acetyltransferase and in turn affect sleep duration. Whether through epigenetics, gene–gene interaction, or isolated polymorphisms, the association of a *CLOCK* variant with sleep duration is indicated here. This finding should in turn support further functional studies.

Our work was supported by the FP6 program EUCLOCK (TR, KVA, MM, MT, TM, and AM), the Dutch Science Foundation (the NWO), The Hersenen Stichting, and the Rosalind Franklin Fellowships of the University of Groningen. The Estonian group was supported by Targeted Financing from Estonian Government SF0180142, European Union through the European Regional Development Fund in the frame of Centre of Excellence in Genomics, and 7 FP Project ECOGENE and OPENGENE. In South Tyrol, the study was supported by the Ministry of Health and Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano, and the South Tyrolean Sparkasse Foundation. We thank the participants in the Estonian Genome project and the South Tyrolean study participants for donating their time and DNA.

All authors report no biomedical financial interests or potential conflicts of interest.

Supplementary material cited in this article is available online.

1. Cirelli C (2006): Cellular consequences of sleep deprivation in the brain. *Sleep Med Rev* 10:307–321.
2. Cirelli C (2006): Sleep disruption, oxidative stress, and aging: New insights from fruit flies. *Proc Natl Acad Sci U S A* 103:13901–13902.
3. Cirelli C, Faraguna U, Tononi G (2006): Changes in brain gene expression after long-term sleep deprivation. *J Neurochem* 98:1632–1645.
4. Gilestro GF, Tononi G, Cirelli C (2009): Widespread changes in synaptic markers as a function of sleep and wakefulness in Drosophila. *Science* 324:109–112.
5. Borbely AA (1998): Processes underlying sleep regulation. *Horm Res* 49:114–117.
6. Daan S, Beersma DG, Borbely AA (1984): Timing of human sleep: Recovery process gated by a circadian pacemaker. *Am J Physiol* 246:R161–R183.
7. Roenneberg T, Merrow M (2007): Entrainment of the human circadian clock. *Cold Spring Harb Symp Quant Biol* 72:293–299.
8. Youngstedt SD, Kripke DF (2004): Long sleep and mortality: Rationale for sleep restriction. *Sleep Med Rev* 8:159–174.
9. Terracciano A, Sanna S, Uda M, Deiana B, Usala G, Busonero F, *et al.* (2008): Genome-wide association scan for five major dimensions of personality [published online ahead of print October 28]. *Mol Psychiatry.*
10. Vincent N, Cox B, Clara I (2009): Are personality dimensions associated with sleep length in a large nationally representative sample? *Compr Psychiatry* 50:158–163.
11. Stickgold R, Hobson JA, Fosse R, Fosse M (2001): Sleep, learning, and dreams: Off-line memory reprocessing. *Science* 294:1052–1057.
12. Van Cauter E, Holmback U, Knutson K, Leproult R, Miller A, Nedeltcheva A, *et al.* (2007): Impact of sleep and sleep loss on neuroendocrine and metabolic function. *Horm Res* 67(suppl 1):2–9.
13. Aeschbach D, Postolache TT, Sher L, Matthews JR, Jackson MA, Wehr TA (2001): Evidence from the waking electroencephalogram that short sleepers live under higher homeostatic sleep pressure than long sleepers. *Neuroscience* 102:493–502.
14. Viola AU, Archer SN, James LM, Groeger JA, Lo JC, Skene DJ, *et al.* (2007): PER3 polymorphism predicts sleep structure and waking performance. *Curr Biol* 17:613–618.
15. Franken P, Dudley CA, Estill SJ, Barakat M, Thomason R, O'Hara BF, *et al.* (2006): NPAS2 as a transcriptional regulator of non-rapid eye movement sleep: Genotype and sex interactions. *Proc Natl Acad Sci U S A* 103:7118–7123.
16. Franken P, Thomason R, Heller HC, O'Hara BF (2007): A non-circadian role for clock-genes in sleep homeostasis: A strain comparison. *BMC Neurosci* 8:87.
17. Partinen M, Kaprio J, Koskenvuo M, Putkonen P, Langinvainio H (1983): Genetic and environmental determination of human sleep. *Sleep* 6:179–185.
18. Tafti M, Franken P (2002): Invited review: Genetic dissection of sleep. *J Appl Physiol* 92:1339–1347.
19. Iwase T, Kajimura N, Uchiyama M, Ebisawa T, Yoshimura K, Kamei Y, *et al.* (2002): Mutation screening of the human clock gene in circadian rhythm sleep disorders. *Psychiatry Res* 109:121–128.
20. King VM, Bentley GE, Follett BK (1997): A direct comparison of photoperiodic time measurement and the circadian system in European starlings and Japanese quail. *J Biol Rhythms* 12:431–442.
21. Naylor E, Bergmann BM, Krauski K, Zee PC, Takahashi JS, Vitaterna MH, *et al.* (2000): The circadian clock mutation alters sleep homeostasis in the mouse. *J Neurosci* 20:8138–8143.
22. Siepka SM, Yoo SH, Park J, Lee C, Takahashi JS (2007): Genetics and neurobiology of circadian clocks in mammals. *Cold Spring Harb Symp Quant Biol* 72:251–259.
23. Allebrandt KV, Roenneberg T (2008): The search for circadian clock components in humans: New perspectives for association studies. *Braz J Med Biol Res* 41:716–721.
24. von Schantz M (2008): Phenotypic effects of genetic variability in human clock genes on circadian and sleep parameters. *J Genet* 87:513–519.
25. Gottlieb DJ, O'Connor GT, Wilk JB (2007): Genome-wide association of sleep and circadian phenotypes. *BMC Med Genet* 8(suppl 1):S9.
26. Pattaro C, Marroni F, Riegler A, Mascalzoni D, Pichler I, Volpato CB, *et al.* (2007): The genetic study of three population microisolates in South Tyrol (MICROS): Study design and epidemiological perspectives. *BMC Med Genet* 8:29.
27. Aulchenko YS, de Koning DJ, Haley C (2007): Genomewide rapid association using mixed model and regression: A fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* 177:577–585.
28. Nelis M, Esko T, Magi R, Zimprich F, Zimprich A, Toncheva D, *et al.* (2009): Genetic structure of Europeans: A view from the North-East. *PLoS ONE* 4:e5472.
29. Kantermann T, Juda M, Merrow M, Roenneberg T (2007): The human circadian clock's seasonal adjustment is disrupted by daylight saving time. *Curr Biol* 17:1996–2000.
30. Barrett JC, Fry B, Maller J, Daly MJ (2005): Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.
31. Miller SA, Dykes DD, Polesky HF (1988): A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215.
32. Storm N, Darnhofer-Patel B, van den Boom D, Rodi CP (2003): MALDI-TOF mass spectrometry-based SNP genotyping. *Methods Mol Biol* 212: 241–262.
33. Heid CA, Stevens J, Livak KJ, Williams PM (1996): Real time quantitative PCR. *Genome Res* 6:986–994.

34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* (2007): PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.
35. Roenneberg T, Kuehnle T, Pramstaller PP, Ricken J, Havel M, Guth A, *et al.* (2004): A marker for the end of adolescence. *Curr Biol* 14:R1038–R1039.
36. Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, Ikeda M, *et al.* (2006): A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat Genet* 38:644–651.
37. Tonetti L, Fabbri M, Natale V (2008): Sex difference in sleep-time preference and sleep need: A cross-sectional survey among Italian pre-adolescents, adolescents, and adults. *Chronobiol Int* 25:745–759.
38. Sowers MF, Zheng H, Kravitz HM, Matthews K, Bromberger JT, Gold EB, *et al.* (2008): Sex steroid hormone profiles are related to sleep measures from polysomnography and the Pittsburgh Sleep quality Index. *Sleep* 31:1339–1349.
39. Greer W, Sandridge AL, Chehabeddine RS (2003): The frequency distribution of age at natural menopause among Saudi Arabian women. *Maturitas* 46:263–272.
40. King DP, Vitaterna MH, Chang AM, Dove WF, Pinto LH, Turek FW, *et al.* (1997): The mouse clock mutation behaves as an antimorph and maps within the W19H deletion, distal of Kit. *Genetics* 146:1049–1060.
41. Iuvone PM, Tosini G, Pozdeyev N, Haque R, Klein DC, Chaurasia SS (2005): Circadian clocks, clock networks, arylalkylamine N-acetyltransferase, and melatonin in the retina. *Prog Retin Eye Res* 24:433–456.
42. Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, *et al.* (2002): Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* 109:307–320.
43. Jeffreys AJ, Kauppi L, Neumann R (2001): Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29:217–222.
44. Serretti A, Benedetti F, Mandelli L, Lorenzi C, Pirovano A, Colombo C, *et al.* (2003): Genetic dissection of psychopathological symptoms: Insomnia in mood disorders and CLOCK gene polymorphism. *Am J Med Genet B Neuropsychiatr Genet* 121B:35–38.
45. Katzenberg D, Young T, Finn L, Lin L, King DP, Takahashi JS, *et al.* (1998): A CLOCK polymorphism associated with human diurnal preference. *Sleep* 21:569–576.
46. Mishima K, Tozawa T, Satoh K, Saitoh H, Mishima Y (2005): The 3111T/C polymorphism of hClock is associated with evening preference and delayed sleep timing in a Japanese population sample. *Am J Med Genet B Neuropsychiatr Genet* 133B:101–104.
47. Sookoian S, Gemma C, Gianotti TF, Burgueno A, Castano G, Pirola CJ (2008): Genetic variants of clock transcription factor are associated with individual susceptibility to obesity. *Am J Clin Nutr* 87:1606–1615.
48. Sookoian S, Castano G, Gemma C, Gianotti TF, Pirola CJ (2007): Common genetic variations in CLOCK transcription factor are associated with nonalcoholic fatty liver disease. *World J Gastroenterol* 13: 4242–4248.
49. Taheri S, Thomas GN (2008): Is sleep duration associated with obesity—where do U stand? *Sleep Med Rev* 12:299–302.
50. Shi J, Wittke-Thompson JK, Badner JA, Hattori E, Potash JB, Willour VL, *et al.* (2008): Clock genes may influence bipolar disorder susceptibility and dysfunctional circadian rhythm. *Am J Med Genet B Neuropsychiatr Genet* 147B:1047–1055.
51. Dahl RE (2004): Regulation of sleep and arousal: Comments on part VII. *Ann N Y Acad Sci* 1021:292–293.
52. Harvey AG, Mullin BC, Hinshaw SP (2006): Sleep and circadian rhythms in children and adolescents with bipolar disorder. *Dev Psychopathol* 18:1147–1168.